



SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: BILLING-MEDEL, PATRICIA A.
COHEN, MAURICE
COLPITTS, TRACEY L.
FRIEDMAN, PAULA N.
HAYDEN, MARK
KLASS, MICHAEL R.
ROBERTS-RAPP, LISA
RUSSELL, JOHN C.
STROUPE, STEPHEN D.
- (ii) TITLE OF THE INVENTION: REAGENTS AND METHODS FOR THE
USEFUL FOR DETECTING DISEASES OF THE GASTROINTESTINAL
TRACT
- (iii) NUMBER OF SEQUENCES: 51
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Abbott Laboratories
 - (B) STREET: 100 Abbott Park Road
 - (C) CITY: Abbott Park
 - (D) STATE: IL
 - (E) COUNTRY: USA
 - (F) ZIP: 60064-3500
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: DOS
 - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/828,856
 - (B) FILING DATE: 31-MAR-1997
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Becker, Cheryl L.
 - (B) REGISTRATION NUMBER: 35,441.
 - (C) REFERENCE/DOCKET NUMBER: 6068.US.P1
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 847/935-1729
 - (B) TELEFAX: 847/938-2623
 - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 241 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

GCCAGGAATA ACTAGAGAGG AACAAATGGGG TTATTCAGAG GTTTTGTTTT CCTCTTAGTT      60
CTGTGCCTGC TGCACCAGTC AAATACTTCC TTCATTAAGC TGAATAATAA TGGCTTTGAA      120
GATATTGTCA TTGTTATAGA TCCTAGTGTG CCAGAAGATG AAAAAATAAT TGAACAAATA      180
GAGGATATGG TGA CTACAGC TTCTACGTAC CTGTTTGAAG CCACAGAAAA AAGATTTTTT      240
T

```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

CTAGAGAGGA ACAATGGGGT TATTCAGAGG TTTTGTTTTT CTCTTAGTTC TGTGCCTGCT      60
GCACCAGTCA AATACTTCCT TCATTAAGCT GAATAATAAT GGCTTTGAAG ATATTGTCAT      120
TGTTATAGAT CCTAGTGTGC CAGAAGATGA AAAAATAATT GAACAAATAG AGGATATGGT      180
GACTACAGCT TCTACGTACC TGTTTGAAGC CACAGAAAA

```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 231 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

TTNTGTAACG AAAAAACCCA TAATCAAGAA GCTCCAAGCC TACAAAACAT AAAGTGCAAT      60
TTTAGAAGTA CATGGGAGGT GATTAGCAAT TCTGAGGATT TTAAAAACAC CATACCCATG      120
GTGACACCAC CTCCTCCACC TGTCTTCTCA TTGCTGAAGA TCAGTCAAAG AATTGTGTGC      180
TTAGTTCTTG ATAAGTCTGG AAGCATGGGG GGTAAGGACC GCCTAAATCG A

```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 237 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGGGGGGTAA	GGACCGCCTA	AATCGAATGA	ATCAAGCAGC	AAAACATTTT	CTGCTGCAGA	60
CTGTTGAAAA	TGGATCCTGG	GTGGGGATGG	TTCACTTTGA	TAGTACTGCC	ACTATTGTAA	120
ATAAGCTAAT	CCAAATAAAA	AGCAGTGATG	AAAGAAACAC	ACTCATGGCA	GGATTACCTA	180
CATATCCTCT	GGGAGGAACT	TCCATCTGCT	CTGGAATTAA	ATATGCATT	CAGGTGA	237

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 216 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTTCCATCTG	CTCTGGAATT	AAATATGCAT	TTCAGGTGAT	TGGAGAGCTA	CATTCCCAAC	60
TCGATGGATC	CGAAGTACTG	CTGCTGACTG	ATGGGGAGGA	TAACACTGCA	AGTTCTTGTA	120
TTGATGAAGT	GAAACAAAGT	GGGGCCATTG	TTCATTTTAT	TGCTTTGGGA	AGAGCTGCTG	180
ATGAAGCAGT	AATAGAGATG	AGCAAGATAA	CAGGAG			216

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 201 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 24
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AATTGATAGT	ACAGTGGGAA	AGGNCACGTT	CTTTCTCATC	ACATGGAACA	GTCTGCCTCC	60
CAGTATTTCT	CTCTGGGATC	CCAGTGGAAC	AATAATGGAA	AATTTCACAG	TGGATGCAAC	120
TTCCAAAATG	GCCTATCTCA	GTATTCCAGG	AACTGCAAAG	GTGGGCACTT	GGGCATACAA	180
TCTTCAAGCC	AAAGCGAACC	C				201

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 241 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCAAATTCTT	CTGTGCCTCC	AATCACAGTG	AATGCTAAAA	TGAATAAGGA	CGTAAACAGT	60
TTCCCCAGCC	CAATGATTGT	TTACGCAGAA	ATTCTACAAG	GATATGTACC	TGTTCTTGGA	120
GCCAAATGTA	CTGCTTTCAT	TGAATCACAG	AATGGACATA	CAGAAGTTTT	GGAACTTTGG	180
GATAATGGTG	CAGGCGCTGA	TTCTTTCAAG	AATGATGGAG	TCTACTCCAG	GTATTTTACA	240
G						241

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 242 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTGCAGGCGC TGATTCTTC AAGAATGATG GAGTCTACTC CAGGTATTTT ACAGCATATA	60
CAGAAAATGG CAGATATAGC TTAAAAGTTC GGGCTCATGG AGGAGCAAAC ACTGCCAGGC	120
TAAAATTACG GCCTCCACTG AATAGAGCCG CGTACATACC AGGCTGGGTA GTGAACGGGG	180
AAATTGAAGC AAACCCGCCA AGACCTGAAA TTGATGAGGA TACTCAGACC ACCTTGAGG	240
AT	242

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 233 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCCGCCAAGA CCTGAAATTG ATGAGGATAC TCAGACCACC TTGGAGGATT TCAGCCGAAC	60
AGCATCCGGA GGTGCATTTG TGGTATCACA AGTCCCAAGC CTTCCCTTGC CTGACCAATA	120
CCCACCAAGT CAAATCACAG ACCTTGATGC CACAGTTCAT GAGGATAAGA TTATTCTTAC	180
ATGGACAGCA CCAGGAGATA ATTTTGATGT TGGAAAAGTT CAACGTTATA TCA	233

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 313 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
(B) LOCATION: 22
(D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
(B) LOCATION: 44
(D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCAATACCCA CCAAGTCAAA TNACAGACCT TGATGCCACA GTTNATGAGG ATAAGATTAT	60
TCTTACATGG ACAGCACCAG GAGATAATTT TGATGTTGGA AAAGTTCAAC GTTATATCAT	120
AAGAATAAGT GCAAGTATTC TTGATCTAAG AGACAGTTTT GATGATGCTC TTCAAGTAAA	180
TACTACTGAT CTGTCACCAA AGGAGGCCAA CTCCAAGGAA AGCTTTGCAT TTAAACCAGA	240
AAATATCTCA GAAGAAAATG CAACCCACAT ATTTATTGCC ATTAAAAGTA TAGATAAAGC	300
ATTTGGCATC AAA	313

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 242 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AAGTATTCTT	GATCTAAGAG	ACAGTTTTGA	TGATGCTCTT	CAAGTAAATA	CTACTGATCT	60
GTCACCAAAG	GAGGCCAACT	CCAAGGAAAG	CTTTGCATTT	AAACCAGAAA	ATATCTCAGA	120
AGAAAATGCA	ACCCACATAT	TTATTGCCAT	TAAAAGTATA	GATAAAAGCA	ATTTGACATC	180
AAAAGTATCC	AACATTGCAC	AAGTAACTTT	GTTTATCCCT	CAAGCAAATC	CTGATGACAT	240
TG						242

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 208 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ANANAATGCA	ACCCACATAT	TTATTGCCAT	TAAAAGTATA	GATAAAAGCA	ATTTGACATC	60
AAAAGTATCC	AACATTGCAC	AAGTAACTTT	GTTTATCCCT	CAAGCAAATC	CTGATGACAT	120
TGATCCTACT	CCTACTCCTA	CTCCTACTCC	TGATAAAAGT	CATAATTCTG	GAGTTAATAT	180
TTCTACGCTG	GTATTGTCTG	TGATTGGG				208

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 201 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CTCCTACTCC	TACTCCTGAT	AAAAGTCATA	ATTCTGGAGT	TAATATTTCT	ACGCTGGTAT	60
TGTCGTGAT	TGGGTCTGTT	GTAATTGTTA	ACTTTATTTT	AAGTACCACC	ATTTGAACCT	120
TAACGAAGAA	AAAAATCTTC	AAGTAGACCT	AGAAGAGAGT	TTAAAAAAC	AAAACAATGT	180
AAGTAAAGGA	TATTTCTGAA	T				201

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 301 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 111
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(ix) FEATURE:

(A) NAME/KEY: base_polymorphism

(B) LOCATION: 244

(D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(ix) FEATURE:

(A) NAME/KEY: base_polymorphism

(B) LOCATION: 284

(D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCTGTTGTAA	TTGTTAACTT	TATTTTAAGT	ACCACCATT	GAACCTTAAC	GAAGAAAAA	60
ATCTTCAAGT	AGACCTAGAA	GAGAGTTTT	AAAAACAAA	CAATGTAAGT	NAAGGATATT	120
TCTGAATCTT	AAAATTCATC	CCATGTGTGA	TCATAAACTC	ATAAAAAATA	TTTAAAGATG	180
TCGGAAAAGG	ATACTTTGAT	TAAATAAAAA	CACTCATGGA	TATGTAAAAA	CTGTCAAGAT	240
TAANATTTAA	TAGTTTCATT	TATTTGTTAT	TTTATTTGTA	AGANATAGTG	ATGAACAAAG	300
A						301

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 229 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAAAAAAATC	TTCAAGTAGA	CCTAGAAGAG	AGTTTTTAAA	AACAAAACAA	TGTAAGTAAA	60
GGATATTTCT	GAATCTTAAA	ATTTCATCCA	TGTGTGATCA	TAACTCATA	AAAAATAATT	120
TAAGATGTCG	GAAAAGGATA	CTTTGATTAA	ATAAAACAC	TCATGGATAT	GTAAAAACTG	180
TCAAGATTAA	AATTTAATAG	TTTCATTTAT	TTGTTATTTT	ATTTGTAAG		229

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3043 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTAGAGAGGA	ACAATGGGGT	TATTCAGAGG	TTTTGTTTTT	CTCTTAGTTC	TGTGCCTGCT	60
GCACCAGTCA	AATACTTCCT	TCATTAAGCT	GAATAATAAT	GGCTTTGAAG	ATATTGTCAT	120
TGTTATAGAT	CCTAGTGTGC	CAGAAGATGA	AAAAATAATT	GAACAAATAG	AGGATATGGT	180
GACTACAGCT	TCTACGTACC	TGTTTGAAGC	CACAGAAAAA	AGATTTTTTT	TCAAAAATGT	240
ATCTATATTA	ATTCCTGAGA	ATTGGAAGGA	AAATCCTCAG	TACAAAAGGC	CAAAACATGA	300
AAACCATAAA	CATGCTGATG	TTATAGTTGC	ACCACCTACA	CTCCCAGGTA	GAGATGAACC	360
ATACACCAAG	CAGTTCACAG	AATGTGGAGA	GAAAGGCGAA	TACATTCAT	TCACCCCTGA	420
CCTTCTACTT	GAAAAAAAAC	AAAATGAATA	TGGACCACCA	GGCAAACGTG	TTGTCCATGA	480
GTGGGCTCAC	CTCCGGTGGG	GAGTGTGTTG	TGAGTACAAT	GAAGATCAGC	CTTTCTACCG	540
TGCTAAGTCA	AAAAAAATCG	AAGCAACAAG	GTGTTCCGCA	GGTATCTCTG	GTAAGAAATAG	600
AGTTTATAAG	TGTCAAGGAG	GCAGCTGTCT	TAGTAGAGCA	TGCAGAATTG	ATTCTACAAC	660
AAAAGTGTAT	GGAAAAGATT	GTCAATTCTT	TCCTGATAAA	GTACAAACAG	AAAAAGCATC	720
CATAATGTTT	ATGCAAAGTA	TTGATTCTGT	TGTTGAATTT	TGTAACGAAA	AAACCCATAA	780
TCAAGAAGCT	CCAAGCCTAC	AAAACATAAA	GTGCAATTTT	AGAAGTACAT	GGGAGGTGAT	840
TAGCAATTCT	GAGGATTTTA	AAAACACCAT	ACCCATGGTG	ACACCACCTC	CTCCACCTGT	900

CTTCTCATTG	CTGAAGATCA	GTCAAAGAAT	TGTGTGCTTA	GTTCTTGATA	AGTCTGGAAG	960
CATGGGGGGT	AAGGACCGCC	TAAATCGAAT	GAATCAAGCA	GCAAAACATT	TCCTGCTGCA	1020
GACTGTTGAA	AATGGATCCT	GGGTGGGGAT	GGTTCACTTT	GATAGTACTG	CCACTATTGT	1080
AAATAAGCTA	ATCCAAATAA	AAAGCAGTGA	TGAAAGAAAC	ACACTCATGG	CAGGATTACC	1140
TACATATCCT	CTGGGAGGAA	CTTCCATCTG	CTCTGGAATT	AAATATGCAT	TTCAGGTGAT	1200
TGGAGAGCTA	CATTCCCAAC	TGCATGGATC	CGAAGTACTG	CTGCTGACTG	ATGGGGAGGA	1260
TAACACTGCA	AGTTCTTGTA	TTGATGAAGT	GAAACAAAGT	GGGGCCATTG	TTCATTTTAT	1320
TGCTTTGGGA	AGAGCTGCTG	ATGAAGCAGT	AATAGAGATG	AGCAAGATAA	CAGGAGGAAG	1380
TCATTTTTAT	GTTTCAGATG	AAGCTCAGAA	CAATGGCCTC	ATTGATGCTT	TTGGGGCTCT	1440
TACATCAGGA	AATACTGATC	TCTCCCAGAA	GTCCCTTCAG	CTCGAAAGTA	AGGGATTAAC	1500
ACTGAATAGT	AATGCCTGGA	TGAACGACAC	TGTCTATAAT	GATAGTACAG	TGGGAAAGGA	1560
CACGTTCTTT	CTCATCACAT	GGAACAGTCT	GCCTCCCAGT	ATTTCTCTCT	GGGATCCAG	1620
TGGAACAATA	ATGGAAAATT	TCACAGTGGA	TGCAACTTCC	AAAATGGCCT	ATCTCAGTAT	1680
TCCAGGAACT	GCAAAGGTGG	GCACTTGGGC	ATACAATCTT	CAAGCCAAAG	CGAACCAGGA	1740
AACATTAAC	ATTACAGTAA	CTTCTCGAGC	AGCAAATCTT	TCTGTGCCTC	CAATCACAGT	1800
GAATGCTAAA	ATGAATAAGG	ACGTAAACAG	TTTCCCAGC	CCAATGATTG	TTTACGCAGA	1860
AATTCTACAA	GGATATGTAC	CTGTTCTTGG	AGCCAATGTG	ACTGCTTTCA	TTGAATCACA	1920
GAATGGACAT	ACAGAAGTTT	TGGAACTTTT	GGATAATGGT	GCAGGCGCTG	ATTCTTTCAA	1980
GAATGATGGA	GTCTACTCCA	GGTATTTTAC	AGCATATACA	GAAATGGCA	GATATAGCTT	2040
AAAAGTTCGG	GCTCATGGAG	GAGCAAACAC	TGCCAGGCTA	AAATTACGGC	CTCCACTGAA	2100
TAGAGCCGCG	TACATACCAG	GCTGGGTAGT	GAACGGGGAA	ATTGAAGCAA	ACCCGCCAAG	2160
ACCTGAAATT	GATGAGGATA	CTCAGACCAC	CTTGGAGGAT	TTCAGCCGAA	CAGCATCCGG	2220
AGGTGCATTT	GTGGTATCAC	AAGTCCCAAG	CCTTCCCTTG	CCTGACCAAT	ACCCACCAAG	2280
TCAAATCACA	GACCTTGATG	CCACAGTTCA	TGAGGATAAG	ATTATTCTTA	CATGGACAGC	2340
ACCAGGAGAT	AATTTTGATG	TTGGAAAAGT	TCAACGTTAT	ATCATAAGAA	TAAGTGCAAG	2400
TATTCTTGAT	CTAAGAGACA	GTTTTGATGA	TGCTCTTCAA	GTAAATACTA	CTGATCTGTC	2460
ACCAAAGGAG	GCCAACTCCA	AGGAAAGCTT	TGCATTTAAA	CCAGAAAATA	TCTCAGAAGA	2520
AAATGCAACC	CACATATTTA	TTGCCATTAA	AAGTATAGAT	AAAAGCAATT	TGACATCAAA	2580
AGTATCCAAC	ATTGCACAAG	TAACCTTTGTT	TATCCCTCAA	GCAAATCCTG	ATGACATTGA	2640
TCCTACTCCT	ACTCCTACTC	CTACTCCTGA	TAAAAGTCAT	AATTCTGGAG	TTAATATTTT	2700
TACGCTGGTA	TTGTCTGTGA	TTGGGTCTGT	TGTAATTGTT	AACCTTTATT	TAAGTACCAC	2760
CATTTGAACC	TTAACGAAGA	AAAAAATCTT	CAAGTAGACC	TAGAAGAGAG	TTTTAAAAAA	2820
CAAAACAATG	TAAGTAAAGG	ATATTTCTGA	ATCTTAAAAT	TCATCCCATG	TGTGATCATA	2880
AACTCATAAA	AATAATTTTA	AGATGTCGGA	AAAGGATACT	TTGATTAAAT	AAAAACACTC	2940
ATGGATATGT	AAAACTGTC	AAGATTAAAA	TTTAATAGTT	TCATTTATTT	GTTATTTTAT	3000
TTGTAAGAAA	TAGTGATGAA	CAAAGATCCT	TTTTCATACT	GAT		3043

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1399 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCAAAATCTT	CTGTGCCTCC	AATCACAGTG	AATGCTAAAA	TGAATAAGGA	CGTAAACAGT	60
TTCCCCAGCC	CAATGATTGT	TTACGCAGAA	ATTCTACAAG	GATATGTACC	TGTTCTTGGA	120
GCCAAATGTGA	CTGCTTTTCAT	TGAATCACAG	AATGGACATA	CAGAAGTTT	GGAACCTTTTG	180
GATAATGGTG	CAGGCGCTGA	TTCTTTCAAG	AATGATGGAG	TCTACTCCAG	GTATTTTACA	240
GCATATACAG	AAAATGGCAG	ATATAGCTTA	AAAGTTCGGG	CTCATGGAGG	AGCAAACACT	300
GCCAGGCTAA	AATTACGGCC	TCCACTGAAT	AGAGCCGCGT	ACATACCAGG	CTGGGTAGTG	360
AACGGGGAAA	TTGAAGCAA	CCCGCCAAGA	CCTGAAATTG	ATGAGGATAC	TCAGACCACC	420
TTGGAGGATT	TCAGCCGAAC	AGCATCCGGA	GGTGCAATTG	TGGTATCACA	AGTCCCAAGC	480
CTTCCCTTGC	CTGACCAATA	CCCACCAAGT	CAAAATCACAG	ACCTTGATGC	CACAGTTTAT	540
GAGGATAAGA	TTATTCTTAC	ATGGACAGCA	CCAGGAGATA	ATTTTGTATG	TGGAAAAGTT	600
CAACGTTATA	TCATAAGAAT	AAGTGCAAGT	ATTCTTGATC	TAAGAGACAG	TTTTGTATGAT	660
GCTCTTCAAG	TAAATACTAC	TGATCTGTCA	CCAAAGGAGG	CCAATCCAA	GGAAAGCTTT	720
GCATTTAAAC	CAGAAAATAT	CTCAGAAGAA	AATGCAACCC	ACATATTTAT	TGCCATTTAA	780
AGTATAGATA	AAAGCAATTT	GACATCAAAA	GTATCCAACA	TTGCACAAGT	AACCTTTGTTT	840
ATCCCTCAAG	CAAATCCTGA	TGACATTGAT	CCTACTCCTA	CTCCTACTCC	TACTCCTGAT	900
AAAAGTCATA	ATTCTGGAGT	TAATATTTCT	ACGCTGGTAT	TGTCTGTGAT	TGGGTCTGTT	960
GTAATTGTGA	ACTTTATTTT	AAGTACCACC	ATTTGAACCT	TAACGAAGAA	AAAAATCTTC	1020

AAGTAGACCT	AGAAGAGAGT	TTTAAAAAAC	AAAACAATGT	AAGTAAAGGA	TATTTCTGAA	1080
TCTTAAATTT	CATCCCATGT	GTGATCATAA	ACTCATAAAA	ATAATTTTAA	GATGTCGGAA	1140
AAGGATACTT	TGATTAAATA	AAAACACTCA	TGGATATGTA	AAAACCTGCA	AGATTAAAAT	1200
TTAATAGTTT	CATTTATTTG	TTATTTTATT	TGTAAGAAAT	AGTGATGAAC	AAAGATCCTT	1260
TTTCATACTG	ATACCTGGTT	GTATATTATT	TGATGCAACA	GTTTTCTGAA	ATGATATTTT	1320
AAATTGCATC	AAGAAATTAA	AATCATCTAT	CTGAGTAGTC	AAAATACAAG	TAAAGGAGAG	1380
CAAATAACA	ACATTTGGA					1399

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3181 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCCAGGAATA	ACTAGAGAGG	AACAATGGGG	TTATTCAGAG	GTTTTGTTTT	CCTCTTAGTT	60
CTGTGCCTGC	TGCACCAAGT	AAATACTTCC	TTCATTAAGC	TGAATAATAA	TGGCTTTTGAA	120
GATATTGTCA	TTGTTATAGA	TCCTAGTGTG	CCAGAAGATG	AAAAAATAAT	TGAACAAATA	180
GAGGATATGG	TGACTACAGC	TTCTACGTAC	CTGTTTGAAG	CCACAGAAAA	AAGATTTTTT	240
TTCAAAAATG	TATCTATATT	AATTCCTGAG	AATTGGAAGG	AAAATCCTCA	GTACAAAAGG	300
CCAAAACATG	AAAACCATAA	ACATGCTGAT	GTTATAGTTG	CACCACCTAC	ACTCCCAGGT	360
AGAGATGAAC	CATACACCAA	GCAGTTCACA	GAATGTGGAG	AGAAAGGCGA	ATACATTCAC	420
TTCACCCCTG	ACCTTCTACT	TGAAAAAATA	CAAAATGAAT	ATGGACCACC	AGGCAAACTG	480
TTTGTCCATG	AGTGGGCTCA	CCTCCGGTGG	GGAGTGTTTG	ATGAGTACAA	TGAAGATCAG	540
CCTTTCTACC	GTGCTAAGTC	AAAAAAAATC	GAAGCAACAA	GGTGTTCGCG	AGGTATCTCT	600
GGTAGAAATA	GAGTTTATAA	GTGTCAAGGA	GGCAGCTGTC	TTAGTAGAGC	ATGCAGAATT	660
GATTCTACAA	CAAAACTGTA	TGGAAAAGAT	TGTCAATTCT	TTCTTGATAA	AGTACAAACA	720
GAAAAAGCAT	CCATAATGTT	TATGCAAAGT	ATTGATTCTG	TTGTTGAATT	TTGTAACGAA	780
AAAACCCATA	ATCAAGGAAG	TCCAAGCCTA	CAAAACATAA	AGTGCAATTT	TAGAAGTACA	840
TGGGAGGTGA	TTAGCAATTC	TGAGGATTTT	AAAAACACCA	TACCCATGGT	GACACCACCT	900
CCTCCACCTG	TCTTCTCATT	GCTGAAGATC	AGTCAAAGAA	TTGTGTGCTT	AGTTCTTGAT	960
AAGTCTGGAA	GCATGGGGGG	TAAGGACCGC	CTAAATCGAA	TGAATCAAGC	AGCAAAAACAT	1020
TTCTTGCTGC	AGACTGTTGA	AAATGGATCC	TGGGTGGGGA	TGGTTCACTT	TGATAGTACT	1080
GCCACTATTG	TAAATAAGCT	AATCCAAATA	AAAAGCAGTG	ATGAAAGAAA	CACACTCATG	1140
GCAGGATTAC	CTACATATCC	TCTGGGAGGA	ACTTCCATCT	GCTCTGGAAT	TAAATATGCA	1200
TTTCAGGTGA	TTGGAGAGCT	ACATTCCCAA	CTCGATGGAT	CCGAAGTACT	GCTGCTGACT	1260
GATGGGGAGG	ATAACACTGC	AAGTTCCTGT	ATTGATGAAG	TGAAACAAAG	TGGGGCCATT	1320
GTTCAATTTA	TTGCTTTGGG	AAGAGCTGCT	GATGAAGCAG	TAATAGAGAT	GAGCAAGATA	1380
ACAGGAGGAA	GTCATTTTAA	TGTTTCAGAT	GAAGCTCAGA	ACAATGGCCT	CATTGATGCT	1440
TTTGGGGCTC	TTACATCAGG	AAATACTGAT	CTCTCCAGA	AGTCCCTTCA	GCTCGAAAGT	1500
AAGGGATTAA	CACTGAATAG	TAATGCCTGG	ATGAACGACA	CTGTCATAAT	TGATAGTACA	1560
GTGGGAAAGG	ACACGTTCTT	TCTCATCACA	TGGAACAGTC	TGCTCCAG	TATTTCTCTC	1620
TGGGATCCCA	GTGGAACAAT	AATGGAAAAT	TTACAGTGG	ATGCAACTTC	CAAAATGGCC	1680
TATCTCAGTA	TTCCAGGAAC	TGCAAAGGTG	GGCACTTGGG	CATACAATCT	TCAAGCCAAA	1740
GCGAACCCAG	AAACATTAAC	TATTACAGTA	ACTTCTCGAG	CAGCAAATTC	TTCTGTGCCT	1800
CCAATCACAG	TGAATGCTAA	AATGAATAAG	GACGTAAACA	GTTTCCCCAG	CCCAATGATT	1860
GTTTACGCAG	AAATTCTACA	AGGATATGTA	CCTGTTCTTG	GAGCCAATGT	GACTGCTTTC	1920
ATTGAATCAC	AGAATGGACA	TACAGAAGTT	TTGGAACTTT	TGGATAATGG	TGCAGGCGCT	1980
GATTCTTTCA	AGAATGATGG	AGTCTACTCC	AGGTATTTTA	CAGCATATAC	AGAAAATGGC	2040
AGATATAGCT	TAAAAGTTTC	GGCTCATGGA	GGAGCAAACA	CTGCCAGGCT	AAAATTACGG	2100
CCTCCACTGA	ATAGAGCCGC	GTACATACCA	GGCTGGGTAG	TGAACGGGGA	AATTGAAGCA	2160
AACCCGCCAA	GACCTGAAAT	TGATGAGGAT	ACTCAGACCA	CCTTGGAGGA	TTTCAGCCGA	2220
ACAGCATCCG	GAGGTGCATT	TGTGGTATCA	CAAGTCCCAA	GCCTTCCCTT	GCCTGACCAA	2280
TACCCACCAA	GTCAAATCAC	AGACCTTGAT	GCCACAGTTC	ATGAGGATAA	GATTATTCTT	2340
ACATGGACAG	CACCAGGAGA	TAATTTTGAT	GTTGGAAAAG	TTCAACGTTA	TATCATAAGA	2400
ATAAGTGCAA	GTATTCTTGA	TCTAAGAGAC	AGTTTGTATG	ATGCTCTTCA	AGTAAATACT	2460
ACTGATCTGT	CACCAAAGGA	GGCCAACTCC	AAGGAAAGCT	TTGCATTTAA	ACCAGAAAAT	2520
ATCTCAGAAG	AAAATGCAAC	CCACATATTT	ATTGCCATTA	AAAGTATAGA	TAAAAGCAAT	2580
TTGACATCAA	AAGTATCCAA	CATTGCACAA	GTAACCTTTG	TTATCCCTCA	AGCAAATCCT	2640
GATGACATTG	ATCCTACTCC	TACTCCTACT	CCTACTCCTG	ATAAAAAGTCA	TAATTTGGA	2700
GTTAATATTT	CTACCTGGT	ATTGTCTGTG	ATTGGGTCTG	TTGTAATTGT	TAACCTTTAT	2760
TTAAGTACCA	CCATTTGAAC	CTTAACGAAG	AAAAAAATCT	TCAAGTAGAC	CTAGAAGAGA	2820

GTTTTAAAAA	ACAAAACAAT	GTAAGTAAAG	GATATTCTG	AATCTTAAAA	TTCATCCCAT	2880
GTGTGATCAT	AAACTCATAA	AAATAATTTT	AAGATGTCGG	AAAAGGATAC	TTTGATTAAA	2940
TAAAAACACT	CATGGATATG	TAAAACTGT	CAAGATTAAA	ATTTAATAGT	TTCATTTATT	3000
TGTTATTTTA	TTTGTAAGAA	ATAGTGATGA	ACAAAGATCC	TTTTTCATAC	TGATACCTGG	3060
TTGTATATTA	TTTGATGCAA	CAGTTTTCTG	AAATGATATT	TCAAATTGCA	TCAAGAAATT	3120
AAAATCATCT	ATCTGAGTAG	TCAAATACA	AGTAAAGGAG	AGCAAATAAA	CAACATTTGG	3180
A						3181

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AGCTCGGAAT	TCCGAGCTTG	GATCCTCTAG	AGCGGCCGCC	GACTAGTGAG	CTCGTCGACC	60
CGGGAATT						68

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AATTAATTCC	CGGGTCGACG	AGCTCACTAG	TCGGCGGCCG	CTCTAGAGGA	TCCAAGCTCG	60
GAATTCCG						68

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AGCGGATAAC	AATTCACAC	AGGA	24
------------	-----------	------	----

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TGTAAACGA	CGGCCAGT	18
-----------	----------	----

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CTGCCAGGCT AAAATTACGG

20

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ATCACAGACC TTGATGCCAC

20

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCTGGTATTG TCTGTGATTG GGTC

24

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CATCAGGATT TGCTTGAGGG

20

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TATTGGTCAG GCAAGGGAAG

20

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GTGTTTGCTC CTCCATGAGC

20

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CAAGTAGAAG GTCAGGGGTG

20

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ATAAGTGTCA AGGAGGCAGC

20

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCAGACTGTT CCATGTGATG

20

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ATGTACCTGT TCTTGAGGCC

20

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ACGTACCTGT TTGAAGCCAC

20

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GGTAAGGACC GCCTAAATCG

20

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GAAGTGAAAC AAAGTGGGGC

20

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TTATCCTCCC CATCAGTCAG

20

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

TCGATTTAGG CGGTCCTTAC

20

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TGTGGCTTCA AACAGGTACG

20

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GGGTAAGGAC CGCCTAAATC GAATG

25

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GAGCCCCAAA AGCATCAATG AGG

23

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 917 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Met	Gly	Leu	Phe	Arg	Gly	Phe	Val	Phe	Leu	Leu	Val	Leu	Cys	Leu	Leu	1	5	10	15
His	Gln	Ser	Asn	Thr	Ser	Phe	Ile	Lys	Leu	Asn	Asn	Asn	Gly	Phe	Glu	20	25	30	
Asp	Ile	Val	Ile	Val	Ile	Asp	Pro	Ser	Val	Pro	Glu	Asp	Glu	Lys	Ile	35	40	45	
Ile	Glu	Gln	Ile	Glu	Asp	Met	Val	Thr	Thr	Ala	Ser	Thr	Tyr	Leu	Phe	50	55	60	
Glu	Ala	Thr	Glu	Lys	Arg	Phe	Phe	Phe	Lys	Asn	Val	Ser	Ile	Leu	Ile	65	70	75	80
Pro	Glu	Asn	Trp	Lys	Glu	Asn	Pro	Gln	Tyr	Lys	Arg	Pro	Lys	His	Glu	85	90	95	
Asn	His	Lys	His	Ala	Asp	Val	Ile	Val	Ala	Pro	Pro	Thr	Leu	Pro	Gly	100	105	110	
Arg	Asp	Glu	Pro	Tyr	Thr	Lys	Gln	Phe	Thr	Glu	Cys	Gly	Glu	Lys	Gly	115	120	125	
Glu	Tyr	Ile	His	Phe	Thr	Pro	Asp	Leu	Leu	Leu	Glu	Lys	Lys	Gln	Asn	130	135	140	
Glu	Tyr	Gly	Pro	Pro	Gly	Lys	Leu	Phe	Val	His	Glu	Trp	Ala	His	Leu	145	150	155	160

Arg Trp Gly Val Phe Asp Glu Tyr Asn Glu Asp Gln Pro Phe Tyr Arg
 165 170 175
 Ala Lys Ser Lys Lys Ile Glu Ala Thr Arg Cys Ser Ala Gly Ile Ser
 180 185 190
 Gly Arg Asn Arg Val Tyr Lys Cys Gln Gly Gly Ser Cys Leu Ser Arg
 195 200 205
 Ala Cys Arg Ile Asp Ser Thr Thr Lys Leu Tyr Gly Lys Asp Cys Gln
 210 215 220
 Phe Phe Pro Asp Lys Val Gln Thr Glu Lys Ala Ser Ile Met Phe Met
 225 230 235 240
 Gln Ser Ile Asp Ser Val Val Glu Phe Cys Asn Glu Lys Thr His Asn
 245 250 255
 Gln Glu Ala Pro Ser Leu Gln Asn Ile Lys Cys Asn Phe Arg Ser Thr
 260 265 270
 Trp Glu Val Ile Ser Asn Ser Glu Asp Phe Lys Asn Thr Ile Pro Met
 275 280 285
 Val Thr Pro Pro Pro Pro Pro Val Phe Ser Leu Leu Lys Ile Ser Gln
 290 295 300
 Arg Ile Val Cys Leu Val Leu Asp Lys Ser Gly Ser Met Gly Gly Lys
 305 310 315 320
 Asp Arg Leu Asn Arg Met Asn Gln Ala Ala Lys His Phe Leu Leu Gln
 325 330 335
 Thr Val Glu Asn Gly Ser Trp Val Gly Met Val His Phe Asp Ser Thr
 340 345 350
 Ala Thr Ile Val Asn Lys Leu Ile Gln Ile Lys Ser Ser Asp Glu Arg
 355 360 365
 Asn Thr Leu Met Ala Gly Leu Pro Thr Tyr Pro Leu Gly Gly Thr Ser
 370 375 380
 Ile Cys Ser Gly Ile Lys Tyr Ala Phe Gln Val Ile Gly Glu Leu His
 385 390 395 400
 Ser Gln Leu Asp Gly Ser Glu Val Leu Leu Thr Asp Gly Glu Asp
 405 410 415
 Asn Thr Ala Ser Ser Cys Ile Asp Glu Val Lys Gln Ser Gly Ala Ile
 420 425 430
 Val His Phe Ile Ala Leu Gly Arg Ala Ala Asp Glu Ala Val Ile Glu
 435 440 445
 Met Ser Lys Ile Thr Gly Gly Ser His Phe Tyr Val Ser Asp Glu Ala
 450 455 460
 Gln Asn Asn Gly Leu Ile Asp Ala Phe Gly Ala Leu Thr Ser Gly Asn
 465 470 475 480
 Thr Asp Leu Ser Gln Lys Ser Leu Gln Leu Glu Ser Lys Gly Leu Thr
 485 490 495
 Leu Asn Ser Asn Ala Trp Met Asn Asp Thr Val Ile Ile Asp Ser Thr
 500 505 510
 Val Gly Lys Asp Thr Phe Phe Leu Ile Thr Trp Asn Ser Leu Pro Pro
 515 520 525
 Ser Ile Ser Leu Trp Asp Pro Ser Gly Thr Ile Met Glu Asn Phe Thr
 530 535 540
 Val Asp Ala Thr Ser Lys Met Ala Tyr Leu Ser Ile Pro Gly Thr Ala
 545 550 555 560
 Lys Val Gly Thr Trp Ala Tyr Asn Leu Gln Ala Lys Ala Asn Pro Glu
 565 570 575
 Thr Leu Thr Ile Thr Val Thr Ser Arg Ala Ala Asn Ser Ser Val Pro
 580 585 590
 Pro Ile Thr Val Asn Ala Lys Met Asn Lys Asp Val Asn Ser Phe Pro
 595 600 605
 Ser Pro Met Ile Val Tyr Ala Glu Ile Leu Gln Gly Tyr Val Pro Val
 610 615 620
 Leu Gly Ala Asn Val Thr Ala Phe Ile Glu Ser Gln Asn Gly His Thr
 625 630 635 640
 Glu Val Leu Glu Leu Leu Asp Asn Gly Ala Gly Ala Asp Ser Phe Lys
 645 650 655
 Asn Asp Gly Val Tyr Ser Arg Tyr Phe Thr Ala Tyr Thr Glu Asn Gly
 660 665 670
 Arg Tyr Ser Leu Lys Val Arg Ala His Gly Gly Ala Asn Thr Ala Arg
 675 680 685

```

Leu Lys Leu Arg Pro Pro Leu Asn Arg Ala Ala Tyr Ile Pro Gly Trp
 690                               695                               700
Val Val Asn Gly Glu Ile Glu Ala Asn Pro Pro Arg Pro Glu Ile Asp
705                               710                               715                               720
Glu Asp Thr Gln Thr Thr Leu Glu Asp Phe Ser Arg Thr Ala Ser Gly
                               725                               730                               735
Gly Ala Phe Val Val Ser Gln Val Pro Ser Leu Pro Leu Pro Asp Gln
                               740                               745                               750
Tyr Pro Pro Ser Gln Ile Thr Asp Leu Asp Ala Thr Val His Glu Asp
                               755                               760                               765
Lys Ile Ile Leu Thr Trp Thr Ala Pro Gly Asp Asn Phe Asp Val Gly
770                               775                               780
Lys Val Gln Arg Tyr Ile Ile Arg Ile Ser Ala Ser Ile Leu Asp Leu
785                               790                               795                               800
Arg Asp Ser Phe Asp Asp Ala Leu Gln Val Asn Thr Thr Asp Leu Ser
                               805                               810                               815
Pro Lys Glu Ala Asn Ser Lys Glu Ser Phe Ala Phe Lys Pro Glu Asn
                               820                               825                               830
Ile Ser Glu Glu Asn Ala Thr His Ile Phe Ile Ala Ile Lys Ser Ile
835                               840                               845
Asp Lys Ser Asn Leu Thr Ser Lys Val Ser Asn Ile Ala Gln Val Thr
850                               855                               860
Leu Phe Ile Pro Gln Ala Asn Pro Asp Asp Ile Asp Pro Thr Pro Thr
865                               870                               875                               880
Pro Thr Pro Thr Pro Asp Lys Ser His Asn Ser Gly Val Asn Ile Ser
                               885                               890                               895
Thr Leu Val Leu Ser Val Ile Gly Ser Val Val Ile Val Asn Phe Ile
900                               905                               910
Leu Ser Thr Thr Ile
915

```

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

```

Ala Asn Ser Ser Val Pro Pro Ile Thr Val Asn Ala Lys Met Asn Lys
 1           5           10           15
Asp Val Asn Ser Phe
20

```

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

```

Asp Asn Gly Ala Gly Ala Asp Ser Phe Lys Asn Asp Gly Val Tyr Ser
 1           5           10           15
Arg Tyr Phe Thr Ala Tyr Thr Glu Asn Gly Arg Tyr Ser Leu Lys
20           25           30

```

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

```

Val Arg Ala His Gly Gly Ala Asn Thr Ala Arg Leu Lys Leu Arg Pro
 1           5           10           15
Pro Leu Asn Arg Ala Ala Tyr Ile
                20

```

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

```

Ser Leu Pro Leu Pro Asp Gln Tyr Pro Pro Ser Gln Ile Thr Asp Leu
 1           5           10           15
Asp Ala Thr Val His Glu Asp Lys Ile Ile Leu Thr Trp Thr Ala Pro
        20           25           30
Gly Asp Asn Phe Asp Val Gly Lys
    35           40

```

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

```

Tyr Asn Glu Asp Gln Pro Phe Tyr Arg Ala Lys Ser Lys Lys Ile Glu
 1           5           10           15
Ala Thr Arg Cys
    20

```

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Leu Ser Arg Ala Cys Arg Ile Asp Ser Thr Thr Lys Leu Tyr Gly Lys
 1 5 10 15
 Asp Cys Gln Phe Phe Pro Asp Lys
 20

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Lys Ser Ser Asp Glu Arg Asn Thr Leu Met Ala Gly Leu Pro Thr Tyr
 1 5 10 15
 Pro Leu Gly Gly
 20

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Glu Ile Asp Glu Asp Thr Gln Thr Thr Leu Glu Asp Phe Ser Arg
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Asp Tyr Lys Asp Asp Asp Asp Lys
 1 5

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Asn Met His Thr Glu His
1 5 10 15
His His His His His
20

We Claim:

1. A purified polynucleotide or fragment thereof derived from a CS193
5 gene, wherein said polynucleotide is capable of selectively hybridizing to the nucleic acid of said CS193 gene and has at least 50% identity to a sequence selected from the group consisting of SEQUENCE ID NOS 1-18, and fragments or complements thereof.
- 10 2. The purified polynucleotide of claim 1, wherein said polynucleotide is produced by recombinant techniques.
3. The purified polynucleotide of claim 1, wherein said polynucleotide is produced by synthetic techniques.
- 15 4. The purified polynucleotide of claim 1, wherein said polynucleotide comprises a sequence encoding at least one CS193 epitope.
5. A recombinant expression system comprising a nucleic acid sequence
20 that includes an open reading frame derived from CS193 operably linked to a control sequence compatible with a desired host, wherein said nucleic acid sequence has at least 50% identity to a sequence selected from the group consisting of SEQUENCE ID NOS 1-18 and fragments or complements thereof.
- 25 6. A cell transfected with the recombinant expression system of claim 5.
7. A CS193 polypeptide having at least 60% identity with an amino acid
sequence selected from the group consisting of SEQUENCE ID NOS 41-49, and
30 fragments thereof.
8. The polypeptide of claim 7, wherein said polypeptide is produced by recombinant techniques.
9. The polypeptide of claim 7, wherein said polypeptide is produced by
35 synthetic techniques.

10. An antibody which specifically binds to at least one CS193 epitope, wherein said CS193 epitope is derived from an amino acid sequence having at least 50% identity to an amino acid sequence selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

5

11. A cell transfected with a nucleic acid sequence encoding at least one CS193 epitope, wherein said nucleic acid sequence is selected from the group consisting of SEQUENCE ID NOS 1-18, and fragments or complements thereof.

10

12. A method for producing a polypeptide comprising at least one CS193 epitope, said method comprising incubating host cells that have been transfected with an expression vector containing a polynucleotide sequence encoding a polypeptide, wherein said polypeptide comprises an amino acid sequence having at least 60% identity with an amino acid sequence selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

15

13. A method for producing antibodies which specifically bind to CS193 antigen, said method comprising administering to an individual an isolated immunogenic polypeptide or fragment thereof in an amount sufficient to elicit an immune response, wherein said immunogenic polypeptide comprises at least one CS193 epitope and has at least 50% identity with a sequence selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

20

14. A method for producing antibodies which specifically bind to CS193 antigen, said method comprising administering to an individual a plasmid comprising a polynucleotide sequence which encodes at least one CS193 epitope derived from a polypeptide having an amino acid sequence selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

25

15. A composition of matter comprising a CS193 polynucleotide or fragment thereof, wherein said polynucleotide has at least 50% identity with a polynucleotide selected from the group consisting of SEQUENCE ID NOS 1-18, and fragments or complements thereof.

30

16. A composition of matter comprising a polypeptide containing at least one CS193 epitope, wherein said polypeptide has at least 60% identity with a sequence

35

selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

5 17. A gene, or a fragment thereof, which codes for a CS193 protein comprising an amino acid sequence that has at least 60% identity with SEQUENCE ID NO 41.

10 18. A gene or fragment thereof comprising DNA having at least 50% identity with SEQUENCE ID NO 16, SEQUENCE ID NO 17, or SEQUENCE ID NO 18.

REAGENTS AND METHODS USEFUL FOR DETECTING
DISEASES OF THE GASTROINTESTINAL TRACT

5

Abstract of the Disclosure

10 A set of contiguous and partially overlapping cDNA sequences and polypeptides
encoded thereby, designated as CS193 and transcribed from GI tract tissue, are
described. These sequences are useful for the detecting, diagnosing, staging,
monitoring, prognosticating, preventing or treating, or determining the predisposition
of an individual to diseases and conditions of the GI tract, such as GI tract cancer. Also
15 provided are antibodies which specifically bind to CS193-encoded polypeptide or
protein, and agonists or inhibitors which prevent action of the tissue-specific CS193
polypeptide, which molecules are useful for the therapeutic treatment of GI tract
diseases, tumors or metastases.



Figure 1-A

>2767646	GCCAGGAATA	ACTAGAGAGG	AACAATGGGG	TTATTCAGAG	GTTTTGTTTT
>774134		CTAGAGAGG	AACAATGGGG	TTATTCAGAG	GTTTTGTTTT
>774134IH		CTAGAGAGG	AACAATGGGG	TTATTCAGAG	GTTTTGTTTT
Consensus	GCCAGGAATA	ACTAGAGAGG	AACAATGGGG	TTATTCAGAG	GTTTTGTTTT
>2767646	CCTCTTAGTT	CTGTGCCTGC	TGCACCAGTC	AAATACTTCC	TTCATTAAGC
>774134	CCTCTTAGTT	CTGTGCCTGC	TGCACCAGTC	AAATACTTCC	TTCATTAAGC
>774134IH	CCTCTTAGTT	CTGTGCCTGC	TGCACCAGTC	AAATACTTCC	TTCATTAAGC
Consensus	CCTCTTAGTT	CTGTGCCTGC	TGCACCAGTC	AAATACTTCC	TTCATTAAGC
>2767646	TGAATAATAA	TGGCTTTGAA	GATATTGTCA	TTGTTATAGA	TCCTAGTGTG
>774134	TGAATAATAA	TGGCTTTGAA	GATATTGTCA	TTGTTATAGA	TCCTAGTGTG
>774134IH	TGAATAATAA	TGGCTTTGAA	GATATTGTCA	TTGTTATAGA	TCCTAGTGTG
Consensus	TGAATAATAA	TGGCTTTGAA	GATATTGTCA	TTGTTATAGA	TCCTAGTGTG
>2767646	CCAGAAGATG	AAAAAATAAT	TGAACAAATA	GAGGATATGG	TGACTACAGC
>774134	CCAGAAGATG	AAAAAATAAT	TGAACAAATA	GAGGATATGG	TGACTACAGC
>774134IH	CCAGAAGATG	AAAAAATAAT	TGAACAAATA	GAGGATATGG	TGACTACAGC
Consensus	CCAGAAGATG	AAAAAATAAT	TGAACAAATA	GAGGATATGG	TGACTACAGC
>2767646	TTCTACGTAC	CTGTTTGAAG	CCACAGAAAA	AAGATTTTTTT	T
>774134	TTCTACGTAC	CTGTTTGAAG	CCACAGAAAA		
>774134IH	TTCTACGTAC	CTGTTTGAAG	CCACAGAAAA	AAGATTTTTTT	TTCAAAAATG
Consensus	TTCTACGTAC	CTGTTTGAAG	CCACAGAAAA	AAGATTTTTTT	TTCAAAAATG
>774134IH	TATCTATATT	AATTCCTGAG	AATTGGAAGG	AAAATCCTCA	GTACAAAAGG
Consensus	TATCTATATT	AATTCCTGAG	AATTGGAAGG	AAAATCCTCA	GTACAAAAGG
>774134IH	CCAAAACATG	AAAACCATAA	ACATGCTGAT	GTTATAGTTG	CACCACCTAC
Consensus	CCAAAACATG	AAAACCATAA	ACATGCTGAT	GTTATAGTTG	CACCACCTAC
>774134IH	ACTCCCAGGT	AGAGATGAAC	CATACACCAA	GCAGTTCACA	GAATGTGGAG
Consensus	ACTCCCAGGT	AGAGATGAAC	CATACACCAA	GCAGTTCACA	GAATGTGGAG
>774134IH	AGAAAGGCGA	ATACATTAC	TTCACCCCTG	ACCTTCTACT	TGAAAAAAAA
Consensus	AGAAAGGCGA	ATACATTAC	TTCACCCCTG	ACCTTCTACT	TGAAAAAAAA
>774134IH	CAAAATGAAT	ATGGACCACC	AGGCAAACTG	TTTGTCCATG	AGTGGGCTCA
Consensus	CAAAATGAAT	ATGGACCACC	AGGCAAACTG	TTTGTCCATG	AGTGGGCTCA
>774134IH	CCTCCGGTGG	GGAGTGTTTG	ATGAGTACAA	TGAAGATCAG	CCTTTCTACC
Consensus	CCTCCGGTGG	GGAGTGTTTG	ATGAGTACAA	TGAAGATCAG	CCTTTCTACC
>774134IH	GTGCTAAGTC	AAAAAAAAATC	GAAGCAACAA	GGTGTTCGCG	AGGTATCTCT
Consensus	GTGCTAAGTC	AAAAAAAAATC	GAAGCAACAA	GGTGTTCGCG	AGGTATCTCT
>774134IH	GGTAGAAATA	GAGTTTATAA	GTGTCAAGGA	GGCAGCTGTC	TTAGTAGAGC
Consensus	GGTAGAAATA	GAGTTTATAA	GTGTCAAGGA	GGCAGCTGTC	TTAGTAGAGC

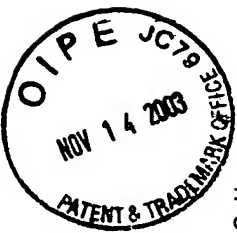


Figure 1-B

>774134IH Consensus	ATGCAGAATT GATTCTACAA CAAAACCTGTA TGGAAAAGAT TGTCAATTCT ATGCAGAATT GATTCTACAA CAAAACCTGTA TGGAAAAGAT TGTCAATTCT
>774134IH Consensus	TTCTTGATAA AGTACAAACA GAAAAAGCAT CCATAATGTT TATGCAAAGT TTCTTGATAA AGTACAAACA GAAAAAGCAT CCATAATGTT TATGCAAAGT
>774134IH >775437 Consensus	ATTGATTCTG TTGTTGAATT TTGTAACGAA AAAACCCATA ATCAAGAAGC TT NTGTAACGAA AAAACCCATA ATCAAGAAGC ATTGATTCTG TTGTTGAATT TTGTAACGAA AAAACCCATA ATCAAGAAGC
>774134IH >775437 Consensus	TCCAAGCCTA CAAAACATAA AGTGCAATTT TAGAAGTACA TGGGAGGTGA TCCAAGCCTA CAAAACATAA AGTGCAATTT TAGAAGTACA TGGGAGGTGA TCCAAGCCTA CAAAACATAA AGTGCAATTT TAGAAGTACA TGGGAGGTGA
>774134IH >775437 Consensus	TTAGCAATTC TGAGGATTTT AAAAAACACCA TACCCATGGT GACACCACCT TTAGCAATTC TGAGGATTTT AAAAAACACCA TACCCATGGT GACACCACCT TTAGCAATTC TGAGGATTTT AAAAAACACCA TACCCATGGT GACACCACCT
>774134IH >775437 Consensus	CCTCCACCTG TCTTCTCATT GCTGAAGATC AGTCAAAGAA TTGTGTGCTT CCTCCACCTG TCTTCTCATT GCTGAAGATC AGTCAAAGAA TTGTGTGCTT CCTCCACCTG TCTTCTCATT GCTGAAGATC AGTCAAAGAA TTGTGTGCTT
>774134IH >775437 >1281329 Consensus	AGTTCTTGAT AAGTCTGGAA GCATGGGGGG TAAGGACCGC CTAAATCGAA AGTTCTTGAT AAGTCTGGAA GCATGGGGGG TAAGGACCGC CTAAATCGA TGGGGGG TAAGGACCGC CTAAATCGAA AGTTCTTGAT AAGTCTGGAA GCATGGGGGG TAAGGACCGC CTAAATCGAA
>774134IH >1281329 Consensus	TGAATCAAGC AGCAAAAACAT TTCCTGCTGC AGACTGTTGA AAATGGATCC TGAATCAAGC AGCAAAAACAT TTCCTGCTGC AGACTGTTGA AAATGGATCC TGAATCAAGC AGCAAAAACAT TTCCTGCTGC AGACTGTTGA AAATGGATCC
>774134IH >1281329 Consensus	TGGGTGGGGA TGGTTCACCT TGATAGTACT GCCACTATTG TAAATAAGCT TGGGTGGGGA TGGTTCACCT TGATAGTACT GCCACTATTG TAAATAAGCT TGGGTGGGGA TGGTTCACCT TGATAGTACT GCCACTATTG TAAATAAGCT
>774134IH >1281329 Consensus	AATCCAAATA AAAAGCAGTG ATGAAAAGAAA CACTCATG GCAGGATTAC AATCCAAATA AAAAGCAGTG ATGAAAAGAAA CACTCATG GCAGGATTAC AATCCAAATA AAAAGCAGTG ATGAAAAGAAA CACTCATG GCAGGATTAC
>774134IH >1281329 >1628677 Consensus	CTACATATCC TCTGGGAGGA ACTTCCATCT GCTCTGGAAT TAAATATGCA CTACATATCC TCTGGGAGGA ACTTCCATCT GCTCTGGAAT TAAATATGCA CTTCCATCT GCTCTGGAAT TAAATATGCA CTACATATCC TCTGGGAGGA ACTTCCATCT GCTCTGGAAT TAAATATGCA
>774134IH >1281329 >1628677 Consensus	TTTCAGGTGA TTGGAGAGCT ACATTCCCAA CTCGATGGAT CCGAAGTACT TTTCAGGTGA TTTCAGGTGA TTGGAGAGCT ACATTCCCAA CTCGATGGAT CCGAAGTACT TTTCAGGTGA TTGGAGAGCT ACATTCCCAA CTCGATGGAT CCGAAGTACT



Figure 1-C

>774134IH	GCTGCTGACT	GATGGGGAGG	ATAAACTGC	AAGTTCTTGT	ATTGATGAAG
>1628677	GCTGCTGACT	GATGGGGAGG	ATAAACTGC	AAGTTCTTGT	ATTGATGAAG
Consensus	GCTGCTGACT	GATGGGGAGG	ATAAACTGC	AAGTTCTTGT	ATTGATGAAG
>774134IH	TGAAACAAAG	TGGGGCCATT	GTTCAATTTA	TTGCTTTGGG	AAGAGCTGCT
>1628677	TGAAACAAAG	TGGGGCCATT	GTTCAATTTA	TTGCTTTGGG	AAGAGCTGCT
Consensus	TGAAACAAAG	TGGGGCCATT	GTTCAATTTA	TTGCTTTGGG	AAGAGCTGCT
>774134IH	GATGAAGCAG	TAATAGAGAT	GAGCAAGATA	ACAGGAGGAA	GTCATTTTTA
>1628677	GATGAAGCAG	TAATAGAGAT	GAGCAAGATA	ACAGGAG	
Consensus	GATGAAGCAG	TAATAGAGAT	GAGCAAGATA	ACAGGAGGAA	GTCATTTTTA
>774134IH	TGTTTCAGAT	GAAGCTCAGA	ACAATGGCCT	CATTGATGCT	TTTGGGGCTC
Consensus	TGTTTCAGAT	GAAGCTCAGA	ACAATGGCCT	CATTGATGCT	TTTGGGGCTC
>774134IH	TTACATCAGG	AAATACTGAT	CTCTCCCAGA	AGTCCCTTCA	GCTCGAAAGT
Consensus	TTACATCAGG	AAATACTGAT	CTCTCCCAGA	AGTCCCTTCA	GCTCGAAAGT
>774134IH	AAGGGATTAA	CACTGAATAG	TAATGCCTGG	ATGAACGACA	CTGTCATAAT
>1286372					AAT
Consensus	AAGGGATTAA	CACTGAATAG	TAATGCCTGG	ATGAACGACA	CTGTCATAAT
>774134IH	TGATAGTACA	GTGGGAAAGG	ACACGTTCTT	TCTCATCACA	TGGAACAGTC
>1286372	TGATAGTACA	GTGGGAAAGG	NCACGTTCTT	TCTCATCACA	TGGAACAGTC
Consensus	TGATAGTACA	GTGGGAAAGG	ACACGTTCTT	TCTCATCACA	TGGAACAGTC
>774134IH	TGCCTCCCAG	TATTTCTCTC	TGGGATCCCA	GTGGAACAAT	AATGGAAAAT
>1286372	TGCCTCCCAG	TATTTCTCTC	TGGGATCCCA	GTGGAACAAT	AATGGAAAAT
Consensus	TGCCTCCCAG	TATTTCTCTC	TGGGATCCCA	GTGGAACAAT	AATGGAAAAT
>774134IH	TTACAGTGG	ATGCAACTTC	CAAAATGGCC	TATCTCAGTA	TTCCAGGAAC
>1286372	TTACAGTGG	ATGCAACTTC	CAAAATGGCC	TATCTCAGTA	TTCCAGGAAC
Consensus	TTACAGTGG	ATGCAACTTC	CAAAATGGCC	TATCTCAGTA	TTCCAGGAAC
>774134IH	TGCAAAGGTG	GGCACTTGGG	CATACAATCT	TCAAGCCAAA	GCGAACCCAG
>1286372	TGCAAAGGTG	GGCACTTGGG	CATACAATCT	TCAAGCCAAA	GCGAACCC
Consensus	TGCAAAGGTG	GGCACTTGGG	CATACAATCT	TCAAGCCAAA	GCGAACCCAG
>774134IH	AAACATTAAC	TATTACAGTA	ACTTCTCGAG	CAGCAAATTC	TTCTGTGCCT
>774419				GCAAATTC	TTCTGTGCCT
>774419IH				GCAAATTC	TTCTGTGCCT
Consensus	AAACATTAAC	TATTACAGTA	ACTTCTCGAG	CAGCAAATTC	TTCTGTGCCT
>774134IH	CCAATCACAG	TGAATGCTAA	AATGAATAAG	GACGTAAACA	GTTTCCCCAG
>774419	CCAATCACAG	TGAATGCTAA	AATGAATAAG	GACGTAAACA	GTTTCCCCAG
>774419IH	CCAATCACAG	TGAATGCTAA	AATGAATAAG	GACGTAAACA	GTTTCCCCAG
Consensus	CCAATCACAG	TGAATGCTAA	AATGAATAAG	GACGTAAACA	GTTTCCCCAG

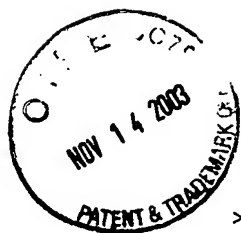


Figure 1-D

>774134IH	CCCAATGATT	GTTTACGCAG	AAATTCTACA	AGGATATGTA	CCTGTTCTTG
>774419	CCCAATGATT	GTTTACGCAG	AAATTCTACA	AGGATATGTA	CCTGTTCTTG
>774419IH	CCCAATGATT	GTTTACGCAG	AAATTCTACA	AGGATATGTA	CCTGTTCTTG
Consensus	CCCAATGATT	GTTTACGCAG	AAATTCTACA	AGGATATGTA	CCTGTTCTTG
>774134IH	GAGCCAATGT	GACTGCTTTC	ATTGAATCAC	AGAATGGACA	TACAGAAGTT
>774419	GAGCCAATGT	GACTGCTTTC	ATTGAATCAC	AGAATGGACA	TACAGAAGTT
>774419IH	GAGCCAATGT	GACTGCTTTC	ATTGAATCAC	AGAATGGACA	TACAGAAGTT
Consensus	GAGCCAATGT	GACTGCTTTC	ATTGAATCAC	AGAATGGACA	TACAGAAGTT
>774134IH	TTGGAAC TTT	TGGATAATGG	TGCAGGCGCT	GATTCTTTCA	AGAATGATGG
>774419	TTGGAAC TTT	TGGATAATGG	TGCAGGCGCT	GATTCTTTCA	AGAATGATGG
>774419IH	TTGGAAC TTT	TGGATAATGG	TGCAGGCGCT	GATTCTTTCA	AGAATGATGG
>3233118		G	TGCAGGCGCT	GATTCTTTCA	AGAATGATGG
Consensus	TTGGAAC TTT	TGGATAATGG	TGCAGGCGCT	GATTCTTTCA	AGAATGATGG
>774134IH	AGTCTACTCC	AGGTATTTTA	CAGCATATAC	AGAAAATGGC	AGATATAGCT
>774419	AGTCTACTCC	AGGTATTTTA	CAG		
>774419IH	AGTCTACTCC	AGGTATTTTA	CAGCATATAC	AGAAAATGGC	AGATATAGCT
>3233118	AGTCTACTCC	AGGTATTTTA	CAGCATATAC	AGAAAATGGC	AGATATAGCT
Consensus	AGTCTACTCC	AGGTATTTTA	CAGCATATAC	AGAAAATGGC	AGATATAGCT
>774134IH	TAAAAGTTCG	GGCTCATGGA	GGAGCAAACA	CTGCCAGGCT	AAAATTACGG
>774419IH	TAAAAGTTCG	GGCTCATGGA	GGAGCAAACA	CTGCCAGGCT	AAAATTACGG
>3233118	TAAAAGTTCG	GGCTCATGGA	GGAGCAAACA	CTGCCAGGCT	AAAATTACGG
Consensus	TAAAAGTTCG	GGCTCATGGA	GGAGCAAACA	CTGCCAGGCT	AAAATTACGG
>774134IH	CCTCCACTGA	ATAGAGCCGC	GTACATACCA	GGCTGGGTAG	TGAACGGGGA
>774419IH	CCTCCACTGA	ATAGAGCCGC	GTACATACCA	GGCTGGGTAG	TGAACGGGGA
>3233118	CCTCCACTGA	ATAGAGCCGC	GTACATACCA	GGCTGGGTAG	TGAACGGGGA
Consensus	CCTCCACTGA	ATAGAGCCGC	GTACATACCA	GGCTGGGTAG	TGAACGGGGA
>774134IH	AATTGAAGCA	AACCCGCCAA	GACCTGAAAT	TGATGAGGAT	ACTCAGACCA
>774419IH	AATTGAAGCA	AACCCGCCAA	GACCTGAAAT	TGATGAGGAT	ACTCAGACCA
>3233118	AATTGAAGCA	AACCCGCCAA	GACCTGAAAT	TGATGAGGAT	ACTCAGACCA
>2733923		CCCGCCAA	GACCTGAAAT	TGATGAGGAT	ACTCAGACCA
Consensus	AATTGAAGCA	AACCCGCCAA	GACCTGAAAT	TGATGAGGAT	ACTCAGACCA
>774134IH	CCTTGAGGGA	TTTCAGCCGA	ACAGCATCCG	GAGGTGCATT	TGTGGTATCA
>774419IH	CCTTGAGGGA	TTTCAGCCGA	ACAGCATCCG	GAGGTGCATT	TGTGGTATCA
>3233118	CCTTGAGGGA	T			
>2733923	CCTTGAGGGA	TTTCAGCCGA	ACAGCATCCG	GAGGTGCATT	TGTGGTATCA
Consensus	CCTTGAGGGA	TTTCAGCCGA	ACAGCATCCG	GAGGTGCATT	TGTGGTATCA
>774134IH	CAAGTCCCAA	GCCTTCCCTT	GCCTGACCAA	TACCCACCAA	GTCAAATCAC
>774419IH	CAAGTCCCAA	GCCTTCCCTT	GCCTGACCAA	TACCCACCAA	GTCAAATCAC
>2733923	CAAGTCCCAA	GCCTTCCCTT	GCCTGACCAA	TACCCACCAA	GTCAAATCAC
>906605			CCAA	TACCCACCAA	GTCAAATNAC
Consensus	CAAGTCCCAA	GCCTTCCCTT	GCCTGACCAA	TACCCACCAA	GTCAAATCAC



Figure 1-E

>774134IH	AGACCTTGAT	GCCACAGTTC	ATGAGGATAA	GATTATTCTT	ACATGGACAG
>774419IH	AGACCTTGAT	GCCACAGTTC	ATGAGGATAA	GATTATTCTT	ACATGGACAG
>2733923	AGACCTTGAT	GCCACAGTTC	ATGAGGATAA	GATTATTCTT	ACATGGACAG
>906605	AGACCTTGAT	GCCACAGTTN	ATGAGGATAA	GATTATTCTT	ACATGGACAG
Consensus	AGACCTTGAT	GCCACAGTTC	ATGAGGATAA	GATTATTCTT	ACATGGACAG
>774134IH	CACCAGGAGA	TAATTTTGAT	GTTGGAAAAG	TTCAACGTTA	TATCATAAGA
>774419IH	CACCAGGAGA	TAATTTTGAT	GTTGGAAAAG	TTCAACGTTA	TATCATAAGA
>2733923	CACCAGGAGA	TAATTTTGAT	GTTGGAAAAG	TTCAACGTTA	TATCA
>906605	CACCAGGAGA	TAATTTTGAT	GTTGGAAAAG	TTCAACGTTA	TATCATAAGA
Consensus	CACCAGGAGA	TAATTTTGAT	GTTGGAAAAG	TTCAACGTTA	TATCATAAGA
>774134IH	ATAAGTGCAA	GTATTCTTGA	TCTAAGAGAC	AGTTTTGATG	ATGCTCTTCA
>774419IH	ATAAGTGCAA	GTATTCTTGA	TCTAAGAGAC	AGTTTTGATG	ATGCTCTTCA
>906605	ATAAGTGCAA	GTATTCTTGA	TCTAAGAGAC	AGTTTTGATG	ATGCTCTTCA
>2771475	AA	GTATTCTTGA	TCTAAGAGAC	AGTTTTGATG	ATGCTCTTCA
Consensus	ATAAGTGCAA	GTATTCTTGA	TCTAAGAGAC	AGTTTTGATG	ATGCTCTTCA
>774134IH	AGTAAATACT	ACTGATCTGT	CACCAAAGGA	GGCCAACCTCC	AAGGAAAGCT
>774419IH	AGTAAATACT	ACTGATCTGT	CACCAAAGGA	GGCCAACCTCC	AAGGAAAGCT
>906605	AGTAAATACT	ACTGATCTGT	CACCAAAGGA	GGCCAACCTCC	AAGGAAAGCT
>2771475	AGTAAATACT	ACTGATCTGT	CACCAAAGGA	GGCCAACCTCC	AAGGAAAGCT
Consensus	AGTAAATACT	ACTGATCTGT	CACCAAAGGA	GGCCAACCTCC	AAGGAAAGCT
>774134IH	TTGCATTTAA	ACCAGAAAAT	ATCTCAGAAG	AAAATGCAAC	CCACATATTT
>774419IH	TTGCATTTAA	ACCAGAAAAT	ATCTCAGAAG	AAAATGCAAC	CCACATATTT
>906605	TTGCATTTAA	ACCAGAAAAT	ATCTCAGAAG	AAAATGCAAC	CCACATATTT
>2771475	TTGCATTTAA	ACCAGAAAAT	ATCTCAGAAG	AAAATGCAAC	CCACATATTT
>1803247			AN	ANAATGCAAC	CCACATATTT
Consensus	TTGCATTTAA	ACCAGAAAAT	ATCTCAGAAG	AAAATGCAAC	CCACATATTT
>774134IH	ATTGCCATTA	AAAGTATAGA	TAAAAGCAAT	TTGACATCAA	AAGTATCCAA
>774419IH	ATTGCCATTA	AAAGTATAGA	TAAAAGCAAT	TTGACATCAA	AAGTATCCAA
>906605	ATTGCCATTA	AAAGTATAGA	TAAA:GCA:T	TTGGCATCAA	A
>2771475	ATTGCCATTA	AAAGTATAGA	TAAAAGCAAT	TTGACATCAA	AAGTATCCAA
>1803247	ATTGCCATTA	AAAGTATAGA	TAAAAGCAAT	TTGACATCAA	AAGTATCCAA
Consensus	ATTGCCATTA	AAAGTATAGA	TAAAAGCAAT	TTGACATCAA	AAGTATCCAA
>774134IH	CATTGCACAA	GTAACTTTGT	TTATCCCTCA	AGCAAATCCT	GATGACATTG
>774419IH	CATTGCACAA	GTAACTTTGT	TTATCCCTCA	AGCAAATCCT	GATGACATTG
>2771475	CATTGCACAA	GTAACTTTGT	TTATCCCTCA	AGCAAATCCT	GATGACATTG
>1803247	CATTGCACAA	GTAACTTTGT	TTATCCCTCA	AGCAAATCCT	GATGACATTG
Consensus	CATTGCACAA	GTAACTTTGT	TTATCCCTCA	AGCAAATCCT	GATGACATTG
>774134IH	ATCCTACTCC	TACTCCTACT	CCTACTCCTG	ATAAAAGTCA	TAATTCTGGA
>774419IH	ATCCTACTCC	TACTCCTACT	CCTACTCCTG	ATAAAAGTCA	TAATTCTGGA
>1803247	ATCCTACTCC	TACTCCTACT	CCTACTCCTG	ATAAAAGTCA	TAATTCTGGA
>1737526		CTCCTACT	CCTACTCCTG	ATAAAAGTCA	TAATTCTGGA
Consensus	ATCCTACTCC	TACTCCTACT	CCTACTCCTG	ATAAAAGTCA	TAATTCTGGA



Figure 1-F

>74134IH	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGGTCTG TTGTAATTGT
>774419IH	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGGTCTG TTGTAATTGT
>1803247	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGG
>1737526	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGGTCTG TTGTAATTGT
>2792957	TCTG TTGTAATTGT
Consensus	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGGTCTG TTGTAATTGT
>774134IH	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAATCT
>774419IH	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAATCT
>1737526	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAATCT
>2792957	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAATCT
>1226186	G AAAAAAATCT
Consensus	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAATCT
>774134IH	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAACAAT GTAAGTAAAG
>774419IH	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAACAAT GTAAGTAAAG
>1737526	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAACAAT GTAAGTAAAG
>2792957	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAACAAT GTAAGTAAAG
>1226186	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAACAAT GTAAGTAAAG
Consensus	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAACAAT GTAAGTAAAG
>774134IH	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAATCATAA
>774419IH	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAATCATAA
>1737526	GATATTTCTG AAT
>2792957	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAATCATAA
>1226186	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAATCATAA
Consensus	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAATCATAA
>774134IH	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAAACT
>774419IH	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAAACT
>2792957	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAAACT
>1226186	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAAACT
Consensus	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAAACT
>774134IH	CATGGATATG TAAAACTGT CAAGATTAAA ATTTAATAGT TTCATTTATT
>774419IH	CATGGATATG TAAAACTGT CAAGATTAAA ATTTAATAGT TTCATTTATT
>2792957	CATGGATATG TAAAACTGT CAAGATTAAAN ATTTAATAGT TTCATTTATT
>1226186	CATGGATATG TAAAACTGT CAAGATTAAA ATTTAATAGT TTCATTTATT
Consensus	CATGGATATG TAAAACTGT CAAGATTAAA ATTTAATAGT TTCATTTATT
>774134IH	TGTTATTTTA TTTGTAAGAA ATAGTGATGA ACAAAGATCC TTTTTCATAC
>774419IH	TGTTATTTTA TTTGTAAGAA ATAGTGATGA ACAAAGATCC TTTTTCATAC
>2792957	TGTTATTTTA TTTGTAAGAN ATAGTGATGA ACAAAGA
>1226186	TGTTATTTTA TTTGTAAG
Consensus	TGTTATTTTA TTTGTAAGAA ATAGTGATGA ACAAAGATCC TTTTTCATAC
>774134IH	TGAT
>774419IH	TGATACCTGG TTGTATATTA TTTGATGCAA CAGTTTTCTG AAATGATATT
Consensus	TGATACCTGG TTGTATATTA TTTGATGCAA CAGTTTTCTG AAATGATATT
>774419IH	TCAAATTGCA TCAAGAAATT AAAATCATCT ATCTGAGTAG TCAAATACA
Consensus	TCAAATTGCA TCAAGAAATT AAAATCATCT ATCTGAGTAG TCAAATACA



Figure 1-G

>774419IH
Consensus

AGTAAAGGAG AGCAAATAAA CAACATTTGG A
AGTAAAGGAG AGCAAATAAA CAACATTTGG A



Figure 2

2767646
774134
774134 IH
775437
1281329
1628677
1285372
774419
774419 IH
3233118
2733923
906605
2771475
1803247
1737526
2792957
1226186

